An invariance principle under the total variation distance

Ivan Nourdin* (Université du Luxembourg)
Guillaume Poly (Université de Rennes 1)

Abstract: Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables, with mean zero and variance one and let $S_n = (X_1 + \ldots + X_n)/\sqrt{n}$. An old and celebrated result of Prohorov [19] asserts that S_n converges in total variation to the standard Gaussian distribution if and only if S_{n_0} has an absolutely continuous component for some integer $n_0 \geq 1$. In the present paper, we give yet another proof of Prohorov's Theorem, but, most importantly, we extend it to a more general situation. Indeed, instead of merely S_n, we consider a sequence of homogoneous polynomials in the X_i. More precisely, we exhibit conditions under which some nonlinear invariance principle, discovered by Rotar [20] and revisited by Mossel, O'Donnel and Oleszkiewicz [16], holds in the total variation topology. There are many works about CLT under various metrics in the literature, but the present one seems to be the first attempt to deal with homogeneous polynomials in the X_i with degree *strictly* greater than one.

Keywords: Convergence in law; convergence in total variation; absolute continuity; invariance principle.

1 Introduction and main results

Let $X_1, X_2, \ldots$ be independent copies of a random variable with mean zero and variance one. According to the central limit theorem, the normalized sums

$$S_n = \frac{X_1 + \ldots + X_n}{\sqrt{n}} \tag{1.1}$$

converge in distribution to the standard normal law $N \sim N(0,1)$. In fact, using, e.g., the second Dini's theorem it is straightforward to prove a stronger result, namely that S_n converges to N in the Kolmogorov distance:

$$\lim_{n\to\infty} d_{Kol}(S_n, N) = 0, \tag{1.2}$$

where $d_{Kol}(U, V) = \sup_{x\in\mathbb{R}} |P(U \leq x) - P(V \leq x)|$.

The total variation distance, which is defined by $d_{TV}(U, V) = \sup_{A\in\mathcal{B}(\mathbb{R})} |P(U \in A) - P(V \in A)|$, is a stronger distance than the Kolmogorov one. In this view, extending the central limit theorem to the total variation topology is an important question. Simple considerations however show that the central limit theorem does not always hold in total variation and that additional assumptions are required. An easy counterexample is when the law of X_1 is discrete. In this case, that of S_n

*I.N. was supported in part by the (french) ANR grant 'Malliavin, Stein and Stochastic Equations with Irregular Coefficients' [ANR-10-BLAN-0121].

is discrete as well and, if we denote by A the discrete support of the sequence S_n, we have both $P(S_n \in A) = 1$ and $P(N \in A) = 0$, which makes impossible the convergence in total variation of S_n towards N.

A complete answer to the problem of whether $d_{TV}(S_n, N) \to 0$ or not is given by an old and celebrated result of Prohorov [19], which provides a simple necessary and sufficient condition to be checked on the law of X_1. To formulate it, we first need to introduce the Lebesgue decomposition (of the distribution) of a random variable. As is well known, each cumulative distribution function (cdf) F can be represented in the form:

$$F(x) = u \int_{-\infty}^{x} g(y)dy + (1-u)G(x), \quad x \in \mathbb{R}, \tag{1.3}$$

where $u \in [0,1]$ is a real number, $g : \mathbb{R} \to [0,\infty)$ is a real function satisfying $\int_{\mathbb{R}} g(y)dy = 1$ and G is a singular cdf (corresponding to a distribution concentrated on a set of zero Lebesgue measure) with $G'(x) = 0$ for almost all x. The real number $u \in [0,1]$ is uniquely determined by F; the density function g is uniquely determined (up to a set of measure zero) provided $u \neq 0$.

Definition 1.1 *When X is a random variable with cdf F, we say that X is* **singular** *if $u = 0$ in (1.3). If $u > 0$, we say that X has an* **absolutely continuous component** *with density g.*

We can now state Prohorov's theorem [19]. A proof will be given in Section 2.4, only to illustrate a possible use of our forthcoming results.

Theorem 1.2 (Prohorov) *One has $d_{TV}(S_n, N) \to 0$ if and only if there exists $n_0 \geqslant 1$ such that the random variable S_{n_0} has an absolutely continuous component.*

Prohorov's theorem has been the starting point of a fruitful line of research around the validity of the central limit theorem under various metrics and the estimation of their associated rates of convergence. Let us only give, here, a small sample of references dealing with this rich and well studied topic. Convergence of densities in L^∞ are studied by Gnedenko and Kolmogorov [10]. On their side, Mamatov and Halikov [12] dealt with the multivariate CLT in total variation. Barron [2] studied the convergence in relative entropy, whereas Shimizu [21] and Johnson and Barron [11] studied the convergence in Fisher information. As far as rates of convergence are concerned, one can quote Mamatov and Sirazdinov [15] for the total variation distance and, more recently, Bobkov, Chistyakov and Götze for bounds in entropy [3], in Fisher information [4] and for Edgeworth-type expansions in the entropic central limit theorem [5]. Finally we mention [7,8] for a variational approach of these issues with some variance bounds.

All the above-mentioned references have in common to 'only' deal with *sums* of independent random variables. In the present paper, in contrast, we will consider highly *non-linear* functionals of independent random variables. It is a much harder framework to work with, precisely because all the nice properties enjoyed by sums of independent variables are no longer valid in this context (in particular, the use of characteristic functions is no longer appropriate).

Let us now turn into the details of the situation we are considering in the present article. Fix a degree of multilinearity $d \geqslant 1$ ($d = 1$ for linear, $d = 2$ for quadratic, etc.) and, for any $n \geqslant 1$, consider a homogeneous polynomial $Q_n : \mathbb{R}^{N_n} \to \mathbb{R}$ of degree d having the form

$$Q_n(\mathbf{x}) = \sum_{i_1,\ldots,i_d=1}^{N_n} a_n(i_1,\ldots,i_d) x_{i_1} \ldots x_{i_d}, \quad \mathbf{x} = (x_1,\ldots,x_{N_n}) \in \mathbb{R}^{N_n}. \tag{1.4}$$

In what follows, otherwise stated we will always assume that the following assumptions take place:

(i) $N_n \to \infty$ as $n \to \infty$;

(ii) [Vanishing on hyperdiagonals] The coefficients $a_n(i_1, \ldots, i_d)$ vanish on hyperdiagonals, that is, $a_n(i_1, \ldots, i_d) = 0$ if $i_j = i_k$ for some $k \neq j \in \{1, \ldots, d\}$;

(iii) [Symmetry] The coefficients $a_n(i_1, \ldots, i_d)$ are symmetric in the indices, that is, one has $a_n(i_{\sigma(1)}, \ldots, i_{\sigma(d)}) = a_n(i_1, \ldots, i_d)$ for all permutation $\sigma \in \mathfrak{S}_d$, all $i_1, \ldots, i_d = 1, \ldots, N_n$ and all $n \geqslant 1$;

(iv) [Normalization] One has $d! \sum_{i_1,\ldots,i_d=1}^{N_n} a_n(i_1, \ldots, i_d)^2 = 1$ for all $n \geqslant 1$;

(v) [Asymptotic negligibility] One has $\lim_{n\to\infty} \max_{1\leqslant i_1 \leqslant n} \sum_{i_2,\ldots,i_d=1}^{N_n} a_n(i_1, \ldots, i_d)^2 = 0$.

Our assumption (iii) is for convenience only and it is of course not a loss of generality. In contrast, property (v) will be important for us. It is indeed the key for having access to the following Theorem 1.3 taken from [16], which will be one of our main tools. The question of an invariance principle for homogeneous polynomials was tackled by Rotar in his paper [20], actually well before the work [16]. In [20], the existence of moments of order strictly greater than 2 is not required (see [20, Proposition 1]). But, instead, one imposes a condition $\Lambda(\alpha)$ on the polynomials (see [20, p.514-515]), which turns out to be slightly more restrictive than (iv) and which is the reason why we have preferred using [16] to prove our results.

Theorem 1.3 (Mossel, O'Donnel, Oleszkiewicz [16]) *Fix a degree $d \geqslant 1$, and let Q_n be a sequence of homogeneous polynomials given by (1.4) and satisfying* (i) *to* (v). *Let $\mathbf{X} = (X_1, X_2, \ldots)$ be a sequence of independent random variables with mean zero and variance one, belonging to $L^{2+\epsilon}(\Omega)$ for some $\epsilon > 0$ (the same ϵ for each X_i). Assume the same for $\mathbf{Y} = (Y_1, Y_2, \ldots)$. Then*

$$\lim_{n\to\infty} d_{Kol}(Q_n(\mathbf{X}), Q_n(\mathbf{Y})) = 0. \tag{1.5}$$

Observe that one can recover (1.2) by simply considering, in (1.5), $d = 1$, $a_n(i) = \frac{1}{\sqrt{n}}$, $1 \leqslant i \leqslant n$ (which satisfies (iv) and (v)) and $Y_1 \sim N(0, 1)$ (which leads to $Q_n(\mathbf{Y}) \sim N(0, 1)$ for any n).

At this stage, it is natural to wonder whether the convergence (1.5) may be strenghtened to the total variation distance as well:

$$\lim_{n\to\infty} d_{TV}(Q_n(\mathbf{X}), Q_n(\mathbf{Y})) = 0. \tag{1.6}$$

Before detailing our answer to this problem, let us first do a quick digression. As we will see in the sequel, the following class of random variables will play the central role in our paper.

Definition 1.4 *For any $p \in]0, 1]$ and $\alpha > 0$, we define the class $\mathcal{C}(p, \alpha)$ as the set of real random variables X satisfying*

$$X \stackrel{\text{law}}{=} \varepsilon(\alpha U + x_0) + (1 - \varepsilon)V, \tag{1.7}$$

where x_0 is a real number, $U \sim \mathcal{U}_{[-1,1]}$ is uniformly distributed on the interval $[-1, 1]$, $\varepsilon \sim \mathcal{B}(p)$ takes its values in $\{0, 1\}$ and satisfies $P(\varepsilon = 1) = p$, and V is any random variable (without specified distribution); moreover U, ε and V are independent.

At first glance, it is not so easy to catch the meaning of (1.7). To help the reader, let us introduce yet another class of random variables.

Definition 1.5 *For any $c, \alpha > 0$, we define the class $\mathcal{G}(c, \alpha)$ as the set of real random variables X having an absolutely continuous component and whose density g (see (1.3)) satisfies $g(x) \geqslant c$ for all $x \in [x_0 - \alpha, x_0 + \alpha]$ and some $x_0 \in \mathbb{R}$.*

When $c, \alpha > 0$ are chosen small enough, the class $\mathcal{G}(c, \alpha)$ is composed of *almost all* real random variables X having an absolutely continuous component. Indeed, the only further restriction we impose is a little condition on its density g, which is e.g. automatically satisfied if there exists at least one x_0 such that g is continuous at x_0.

The following result compares the two classes $\mathcal{C}(p, \alpha)$ and $\mathcal{G}(c, \alpha)$. Roughly speaking, it asserts that the class of random variables with an absolutely continous component (that is, exactly the kind of random variables appearing in Prohorov's Theorem 1.2) coincides with $\cup_{p\in]0,1],\alpha>0} \mathcal{C}_{p,\alpha}$. Observe also that $\mathcal{G}(c, \alpha)$ is not empty if and only if $2c\alpha \leqslant 1$.

Proposition 1.6 *Fix $c, \alpha > 0$ and $p \in]0, 1]$. One has $\mathcal{G}(c, \alpha) \subset \mathcal{C}(2c\alpha, \alpha)$. Moreover, any random variable belonging to $\mathcal{C}(p, \alpha)$ has an absolutely continuous part.*

In Lemma 2.2 below, we will state two further important properties of $\mathcal{C}(p, \alpha)$. Firstly, the sum of two independent random variables having an absolutely continuous component belongs to $\cup_{c,\alpha>0} \mathcal{G}(c, \alpha) \subset \cup_{p\in]0,1],\alpha>0} \mathcal{C}(p, \alpha)$. Secondly, if $0 < q \leqslant p \leqslant 1$ and $0 < \beta \leqslant \alpha$, then $\mathcal{C}(p, \alpha) \subset \mathcal{C}(q\beta/\alpha, \beta)$.

Now $\mathcal{C}(p, \alpha)$ has been introduced and is hopefully better understood, let us give a label to the set of sequences of independent and normalized random variables we will deal with throughout the sequel.

Definition 1.7 *Let $\alpha > 0$, $p \in]0, 1]$ and $\epsilon > 0$. A sequence $\mathbf{X} = (X_1, X_2, \ldots)$ of random variables is said to belong to $\mathcal{D}(\alpha, p, 2 + \epsilon)$ if the X_i are independent, if they satisfy $\sup_i E|X_i|^{2+\epsilon} < \infty$ and if, for each i, $E[X_i] = 0$, $E[X_i^2] = 1$ and $X_i \in \mathcal{C}(p, \alpha)$.*

We are now in a position to state the main result of the present paper.

Theorem 1.8 *Fix a degree $d \geqslant 1$, and let Q_n be a sequence of homogeneous polynomials given by (1.4) and satisfying the assumptions* (i) *to* (v)*. Let $\mathbf{X}$ and $\mathbf{Y}$ belong to $\mathcal{D}(\alpha, p, 2 + \epsilon)$ for some $\epsilon, \alpha > 0$ and some $p \in]0, 1]$. Then (1.6) holds true.*

A noticeable corollary of Theorem 1.8 is a new proof of Prohorov's Theorem 1.2. We refer the reader to Section 2.4 for the details. Let us just stress, here, that we are indeed able to recover Theorem 1.2 in its *full* generality. This is because we are dealing with the case $d = 1$. Indeed, as we will see in Section 2.4, the fact that $d = 1$ will enable us to not suppose that $\mathbf{X}$ belongs to some $\mathcal{D}(\alpha, p, 2 + \epsilon)$; simply assuming that the X_i's are i.i.d. and square integrable will be enough to conclude.

Another corollary of Theorem 1.8 is the following result.

Corollary 1.9 *Fix a degree $d \geqslant 1$, and let Q_n be a sequence of homogeneous polynomials given by (1.4) and satisfying the assumptions* (i) *to* (v). *Let* $\mathbf{X}$ *belong to* $\mathcal{D}(\alpha, p, 2+\epsilon)$ *for some* $\epsilon, \alpha > 0$ *and some* $p \in]0,1]$. *If* $Q_n(\mathbf{X})$ *converges in law to* W, *then* W *has a density and* $Q_n(\mathbf{X})$ *converges to* W *in total variation.*

Corollary 1.9 would be clearly wrong without assuming (v). For a counterexample, consider, e.g., $Q_n(\mathbf{x}) = x_1$, $n \geqslant 1$ with X_1 singular.

Yet another interesting consequence of Theorem 1.8 is provided by the next theorem.

Theorem 1.10 *Assume that* $d \geq 2$ *(pay attention that* $d = 1$ *is prohibited here) and let* N_n *satisfy* (i). *Let* $\{a_n(i_1, \ldots, i_d)\}_{1 \leqslant i_1,\ldots,i_d \leqslant N_n}$ *be an array of real numbers satisfying* (ii) *to* (iv). *(We need* <u>*not*</u> *suppose* (v).*) Consider* $Q_n(\mathbf{x})$ *defined by (1.4). Let* $\mathbf{X} = (X_1, X_2, \ldots)$ *be a sequence of independent and identically distributed random variables with* $3 \leqslant E[X_1^4] < \infty$, $E[X_1] = E[X_1^3] = 0$ *and* $E[X_1^2] = 1$. *Finally, let* $N \sim N(0,1)$. *As* $n \to \infty$, *if*

(a) $E\left[Q_n(\mathbf{X})^4\right] \to E[N^4] = 3$

then

(b) for all $\mathbf{Z} = (Z_1, Z_2, \ldots)$ *belonging to* $\mathcal{D}(\alpha, p, 2+\epsilon)$ *for some* $\epsilon, \alpha > 0$ *and* $p \in]0,1]$, *we have*

$$d_{TV}\left(Q_n(\mathbf{Z}), N\right) \to 0.$$

The rest of our paper is organised as follows. In Section 2 we prove all the results that are stated in this Introduction, except Theorem 1.8; in particular, Section 2.4 contains our new proof of Prohorov's Theorem 1.2. Finally, the proof of our main result, namely Theorem 1.8, is given in Section 3.

2 Proofs of all stated results except Theorem 1.8

2.1 Some useful lemmas

The following lemma will be used several times in the sequel.

Lemma 2.1 *Fix* $q \in [0,1]$, *and let* Y, Z *be two random variables satisfying* $E[f(Y)] \geqslant qE[f(Z)]$ *for all positive bounded function* f. *Then there exists two independent random variables* W *and* $\zeta \sim \mathcal{B}(q)$, *independent from* Z, *such that*

$$Y \overset{\text{law}}{=} \zeta Z + (1-\zeta)W. \tag{2.8}$$

Proof of Lemma 2.1. Our assumption ensures that the linear form $f \mapsto E[f(Y)] - qE[f(Z)]$ is positive. From the Riesz representation theorem, one deduces the existence of a positive finite Radon measure ν such that

$$E[f(Y)] = qE[f(Z)] + \int_{\mathbb{R}} f(x)d\nu(x). \tag{2.9}$$

Choosing $f \equiv 1$ in (2.9) gives $\nu(\mathbb{R}) = 1 - q$. If $\nu(\mathbb{R}) = 0$ then $q = 1$ and the proof of (2.8) is established. Otherwise, $\nu(\mathbb{R}) > 0$ and one can consider $W \sim \frac{1}{\nu(\mathbb{R})}d\nu(x)$, implying in turn (2.8).

□

In the following lemma, we gather useful properties of the classes $\mathcal{C}(p,\alpha)$ and $\mathcal{G}(c,\alpha)$.

Lemma 2.2 *The following properties take place.*

1. *If $0 < q \leqslant p \leqslant 1$ and if $0 < \beta \leqslant \alpha$, then $\mathcal{C}(p,\alpha) \subset \mathcal{C}(q\beta/\alpha,\beta)$. In particular, $\mathcal{C}(p,\alpha) \subset \mathcal{C}(q,\alpha)$.*
2. *If X and Y both have an absolutely continuous component and if X is independent from Y, then there exists $c,\alpha > 0$ such that $X+Y \in \mathcal{G}(c,\alpha)$.*
3. *If X belongs to $\mathcal{C}(p,\alpha)$ with $\alpha > 0$ and $p \in]0,1]$ and if Y is any random variable independent from X, then $X+Y$ belongs to $\mathcal{C}(q,\beta)$ for some $\beta > 0$ and $q \in]0,1]$.*
4. *If $a \neq 0$ and b are two real numbers and if X belongs to $\mathcal{C}(p,\alpha)$ with $\alpha > 0$ and $p \in]0,1]$, then $aX+b \in \mathcal{C}(p,|a|\alpha)$.*

Proof. 1. Fix $0 < q \leqslant p \leqslant 1$ and $0 < \beta \leqslant \alpha$, and consider $X \in \mathcal{C}(p,\alpha)$. According to (1.7), we have, for any positive f,

$$\begin{aligned} E[f(X)] &= p\int_{\mathbb{R}} f(x)\frac{1}{2\alpha}\mathbf{1}_{[x_0-\alpha,x_0+\alpha]}(x)dx + (1-p)E[f(V)] \\ &\geqslant \frac{q\beta}{\alpha}\int_{\mathbb{R}} f(x)\frac{1}{2\beta}\mathbf{1}_{[x_0-\beta,x_0+\beta]}(x)dx. \end{aligned}$$

The conclusion follows from Lemma 2.1.
2. Consider the decomposition (1.3) of the cdf F of X. This defines $u \in]0,1]$ and $g : \mathbb{R}_+ \to \mathbb{R}$ unambiguously. Define similarly $v \in]0,1]$ and $h : \mathbb{R}_+ \to \mathbb{R}$ for Y. For any Borel set A, one has $P(X \in A) \geqslant u\int_A g(x)dx$ and the same for Y. We deduce

$$P(X+Y \in A) \geqslant uv\int_A (g\star h)(x)dx,$$

with $\star$ denoting the usual convolution. Besides, $g\star h = \lim_{M\to\infty} g\star\inf(h,M)$ and the limit is increasing by positivity of g. Finally, we note that, since $g \in L^1$ and $\inf(h,M) \in L^\infty$, the convolution $g\star\inf(h,M)$ is continuous. Let $x_0 \in \mathbb{R}$ and $M > 0$ be such that $(g\star\inf(h,M))(x_0) > 0$. (Such a pair (x_0,M) necessarily exists, otherwise we would have $g\star h \equiv 0$ by taking the large M limit.) By continuity, there exists $c > 0$ and $\alpha > 0$ such that, for any $x \in]x_0-\alpha,x_0+\alpha[$, $(g\star h)(x) \geqslant (g\star\inf(h,M))(x) \geq c$. That is, $X+Y$ belongs to $\mathcal{G}(c,\alpha)$.
3. We have $X \overset{\text{law}}{=} \varepsilon(\alpha U + x_0) + (1-\varepsilon)V$, with $x_0 \in \mathbb{R}$ a real number, and $U \sim \mathcal{U}_{[-1,1]}$, $\varepsilon \sim \mathcal{B}(p)$ and V (with no specified distribution) three independent random variables. On the other hand, one can write $Y \overset{\text{law}}{=} \varepsilon Y + (1-\varepsilon)Z$, with Z having the same law as Y and being independent from Y, U, ε, V. Thus,

$$X+Y \overset{\text{law}}{=} \varepsilon(\alpha U + x_0 + Y) + (1-\varepsilon)(V+Z).$$

The random variable $\alpha U + x_0 + Y$ has a density g given by

$$g(v) = \int_{\mathbb{R}} \frac{1}{2\alpha}\mathbf{1}_{[x_0-\alpha,x_0+\alpha]}(v-y)dP_Y(y) = \frac{1}{2\alpha}P(Y \in [v-x_0-\alpha, v-x_0+\alpha]).$$

As a matter of fact, g is a regulated function, since it is the difference of two increasing functions. In particular, the set $\mathcal{E}$ of its discontinuous points is countable. As a consequence, $\mathrm{Leb}(\mathcal{E}) = 0$, implying in turn $1 = \int_{\mathbb{R}} g(v)dv = \int_{\mathbb{R}\setminus\mathcal{E}} g(v)dv$, so that there exists $x_1 \notin \mathcal{E}$ satisfying $g(x_1) > 0$. Since g is continuous at x_1, there exists $r > 0$ such that $g(v) \geqslant \frac{1}{2}g(x_1)$ for all $v \in [x_1 - r, x_1 + r]$. By Lemma 2.1, it comes that

$$\alpha U + x_0 + Y \overset{\text{law}}{=} \eta(rU + x_1) + (1 - \eta)T,$$

where $\eta \sim \mathcal{B}(p')$ for some $p' \in]0, 1]$, $U \sim \mathcal{U}_{[-1,1]}$ and T are independent. Hence

$$X + Y \overset{\text{law}}{=} \varepsilon\eta(rU + x_1) + \varepsilon(1 - \eta)T + (1 - \varepsilon)(V + Z).$$

As a result, for any bounded positive function,

$$\begin{aligned} E[f(X + Y)] &= pp'E[f(rU + x_1)] + p(1 - p')E[f(T)] + (1 - p)E[f(V + Z)] \\ &\geqslant pp'E[f(rU + x_1)]. \end{aligned}$$

Finally, one deduces that $X + Y$ belongs to $\mathcal{C}(pp', r)$ by Lemma 2.1.
4. Obvious.

□

2.2 Proof of Proposition 1.6

Let X be an element of $\mathcal{G}(c, \alpha)$. Let F denote its cdf, and consider g, G and u as in (1.3). Let $Y \sim g(x)dx$, $Z \sim dG(x)$ and $\eta \sim \mathcal{B}(u)$ be three independent random variables. Then $X \overset{\text{law}}{=} \eta Y + (1 - \eta)Z$. Using the assumption made on g, one obtains, for any positive function f,

$$E[f(Y)] \geqslant 2c\alpha\, E\left[f\left(\alpha U + x_0\right)\right].$$

Observe that $0 < 2c\alpha \leqslant 1$ necessarily. We then deduce that $X \in \mathcal{C}(2c\alpha, \alpha)$ from Lemma 2.1.

Consider now a random variable X belonging to $\mathcal{C}(p, \alpha)$. We have, for any positive function f and according to the decomposition (1.7),

$$E[f(X)] = pE[f(\alpha U + x_0)] + (1 - p)E[f(V)]. \tag{2.10}$$

Let us consider the Lebesgue decomposition (u, g, G) of V, see (1.3):

$$E[f(V)] = u \int_{\mathbb{R}} f(x)g(x)dx + (1 - u) \int_{\mathbb{R}} f(x)dG(x).$$

Plugging into (2.10) yields

$$E[f(X)] = \int_{\mathbb{R}} f(x) \left\{ \frac{p}{2\alpha} \mathbf{1}_{[x_0-\alpha, x_0+\alpha]}(x) + (1 - p)u\, g(x) \right\} dx + (1 - p)(1 - u) \int_{\mathbb{R}} f(x)dG(x),$$

from which we deduce that X has an absolutely continuous component.

□

2.3 Proof of Corollary 1.9

Let the assumption of Corollary 1.9 prevail and consider a sequence $\mathbf{G} = (G_1, G_2, \dots)$ composed of independent copies of a standard Gaussian random variable. By the Mossel, O'Donnel, Oleszkiewicz's invariance principle (Theorem 1.3), one has that $Q_n(\mathbf{G})$ converges in law to W. We shall use the next result taken from [18, Theorem 3.1], that we restate for the reader convenience.

Lemma 2.3 *Let Q_n be any sequence of multivariate polynomials (non necessarily homogeneous) of fixed degree d. Assume that $Q_n(\boldsymbol{G})$ converges in distribution towards W and that $Var(W) > 0$. Then, W is absolutely continuous with respect to the Lebesgue measure and*

$$d_{TV}(Q_n(\boldsymbol{G}), W) \to 0.$$

Let us go back to the proof of Corollary 1.9. Thanks to assumption (iv), it is obvious that $\mathrm{Var}(W) > 0$. But we also know that $d_{TV}(Q_n(\mathbf{X}), Q_n(\mathbf{G})) \to 0$ by our Theorem 1.8. As a result,

$$d_{TV}(Q_n(\mathbf{X}), W) \leqslant d_{TV}(Q_n(\mathbf{X}), Q_n(\mathbf{G})) + d_{TV}(Q_n(\mathbf{G}), W) \to 0 \quad \text{as } n \to \infty.$$

This concludes the proof of Corollary 1.9. □

2.4 Proof of Theorem 1.2

We would like to use our Corollary 1.9 in order to prove Theorem 1.2. But the problem we face is that $\mathbf{X}$ is *not* assumed to belong to some $\mathcal{D}(\alpha, p, 2+\epsilon)$ as is required in the statement of Theorem 1.2. To overcome this difficulty, we shall first make use of Lemma 2.2.

Let the assumptions and notation of Theorem 1.2 prevail. First, if S_n is singular, then there exists a Borel set A such that $P(N \in A) = 0$ and $P(S_n \in A) = 1$; in particular, $d_{TV}(S_n, N) = 1$. Hence, if S_n is singular for all n, then the convergence of S_n to N cannot take place in total variation.

Now, assume the existence of n_0 such that S_{n_0} has an absolutely continuous component. The convergence in total variation of S_n to N will obviously follow from the fact that, for all $k \in \{0, \dots, 2n_0 - 1\}$,

$$\lim_{n\to\infty} d_{TV}(S_{2n_0 n+k}, N) = 0. \tag{2.11}$$

So, fix $k \in \{0, \dots, 2n_0 - 1\}$ and let us prove (2.11). We consider a sequence $(Y_2, Y_3, \dots)$ of independent copies of S_{2n_0}. By Lemma 2.2 (points 2 and 4), observe that each Y_i belongs to $\mathcal{C}(p, \alpha)$, for some $p \in]0,1]$ and $\alpha > 0$ (the same p and the same α for all $i \geqslant 2$); also, we have $E[Y_i] = 0$ and $E[Y_i^2] = 1$. On the other hand, let Y_1 be independent of $Y_2, Y_3, \dots$ and have the same law than S_{2n_0+k}. By Lemma 2.2 (points 3 and 4), Y_1 belongs to $\mathcal{C}(q, \beta)$ for some $q \in]0,1]$ and $\beta > 0$; also, we have $E[Y_1] = 0$ and $E[Y_1^2] = 1$. In fact, thanks to Lemma 2.2 (point 1), one may and will choose the same p and the same α for each Y_i, without making a difference between $i = 1$ and $i \geqslant 2$.

Bearing all the previous notation in mind, we can write

$$S_n \overset{\text{law}}{=} \sqrt{\frac{2n_0 + k}{n}}\, Y_1 + \sqrt{\frac{2n_0}{n}} \sum_{k=2}^{n} Y_k.$$

If Theorem 1.8 were true without assuming $E|Y_1|^q < \infty$ for some $q > 2$, the desired convergence (2.11) would then be a direct consequence of Theorem 1.8 applied with $d = 1$, $\mathbf{X} = \mathbf{G} = (G_1, G_2, \ldots)$ composed of independent $N(0,1)$ variables, and $a_n(1) = \sqrt{\frac{2n_0+k}{n}}$ and $a_n(i) = \sqrt{\frac{2n_0}{n}}$, $i = 2, \ldots, n$. In fact, a careful inspection of the forthcoming proof of Theorem 1.8 would reveal that, when $d = 1$, Theorem 1.8 holds true under the sole assumption that $E|Y_1|^2 < \infty$; this is because, in Step 7 and when $d = 1$, one can rely on the usual CLT instead of Theorem 1.3.

Since, when $d = 1$, the conclusion of Theorem 1.8 appears to be true by only assuming $E|Y_1|^2 < \infty$ (and not $E|Y_1|^q < \infty$ for some $q > 2$), the proof of Theorem 1.2 is concluded. □

2.5 Proof of Theorem 1.10

Assume (a) and let us prove (b). The proof is done is several steps.

Step 1. Let $\mathbf{G} = (G_1, G_2, \ldots)$ stand for a sequence of independent $N(0,1)$ random variables. From [17, Formula 3.1] and our assumptions on the moments of $\mathbf{X}$ one has that

$$E[Q_n(\mathbf{X})^4] \geqslant E[Q_n(\mathbf{G})^4] \geqslant 3.$$

Thus, we deduce from (a) that $E[Q_n(\mathbf{G})^4] \to 3$.

Step 2. One has, according to [13, (11.4.7) and (11.4.8) pp. 192-193]:

$$\max_{1\leqslant i_1 \leqslant n} \sum_{i_2,\ldots,i_d=1}^{N_n} a_n(i_1, \ldots, i_d)^2 \leqslant \frac{1}{dd!}\sqrt{E[Q_n(\mathbf{G})^4] - 3}.$$

As a result, hypothesis (v) for the array $\{a_n(i_1, \ldots, i_d)\}$ turns out to be automatically satisfied.

Step 3 (conclusion). Let $\mathbf{Z} = (Z_1, Z_2, \ldots)$ be as in (b). Theorem 1.8 with $\mathbf{Y} := \mathbf{G}$ implies that $\lim_{n\to\infty} d_{TV}(Q_n(\mathbf{Z}), Q_n(\mathbf{G})) = 0$. On the other hand, thanks to (a) and [13, Theorem 5.2.6] one has that $\lim_{n\to\infty} d_{TV}(Q_n(\mathbf{G}), N) = 0$. Finally, the triangle inequality for d_{TV} implies the desired conclusion (b). □

3 Proof of Theorem 1.8

Let the assumptions and notation of Theorem 1.8. Without loss of generality, for simplicity we assume that $N_n = n$.

During all the proof, the notation $\sum_{i_1,\ldots,i_d=1}^{n}$ is short-hand for $\sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n}$.

The proof is divided into several steps.

Step 1. In the definition of $Q_n(\mathbf{X})$ one may and will replace each X_i by $\varepsilon_i(\alpha U_i + x_i) + (1-\varepsilon_i)V_i$, where $\mathbf{e} = (\varepsilon_1, \varepsilon_2, \ldots)$ is a sequence of independent Bernoulli random variables ($\varepsilon_i \sim \mathcal{B}(p_i)$), $\mathbf{U} = (U_1, U_2, \ldots)$ is a sequence of independent $[-1, 1]$-uniformly distributed random variables and $\mathbf{V} = (V_1, V_2, \ldots)$ is a sequence of independent random variables; moreover, $\mathbf{e}$, $\mathbf{U}$ and $\mathbf{V}$ are independent. That is,

$$Q_n(\mathbf{X}) = \sum_{i_1,\ldots,i_d=1}^{n} a_n(i_1,\ldots,i_d)\{\varepsilon_{i_1}(\alpha U_{i_1}+x_{i_1})+(1-\varepsilon_{i_1})V_{i_1}\}\ldots\{\varepsilon_{i_d}(\alpha U_{i_d}+x_{i_d})+(1-\varepsilon_{i_d})V_{i_d}\}.$$

Now, let us expand everything, and then rewrite $Q_n(\mathbf{X})$ as a polynomial in the U_i. We obtain

$$Q_n(\mathbf{X}) = A_n + B_n + C_n,$$

where

$$\begin{aligned}
A_n &= \sum_{i_1,\ldots,i_d=1}^{n} a_n(i_1,\ldots,i_d)\alpha^d \varepsilon_{i_1}\ldots\varepsilon_{i_d} U_{i_1}\ldots U_{i_d} \\
B_n &= Q_n(\mathbf{X}) - A_n - C_n \\
C_n &= \sum_{i_1,\ldots,i_d=1}^{n} a_n(i_1,\ldots,i_d)\{\varepsilon_{i_1}x_{i_1} + (1-\varepsilon_{i_1})V_{i_1}\}\ldots\{\varepsilon_{i_d}x_{i_d} + (1-\varepsilon_{i_d})V_{i_d}\}
\end{aligned}$$

satisfy

$$E[A_nB_n|\mathbf{e},\mathbf{V}] = E[A_nC_n|\mathbf{e},\mathbf{V}] = E[B_nC_n|\mathbf{e},\mathbf{V}] = E[A_n|\mathbf{e},\mathbf{V}] = E[B_n|\mathbf{e},\mathbf{V}] = 0. \qquad (3.12)$$

Indeed, when seeing $Q_n(\mathbf{X})$ as a multivariate polynomials in the sequence $\{U_i\}_{i\geq 1}$, A_n is the term of maximal degree (i.e. d), C_n is the constant term and B_n is the sum of the remaining terms. Thus, the orthogonality relations (3.12) come from the fact that two homogeneous (see (ii)) polynomials in $\{U_i\}_{i\geq 1}$ having different degrees are orthogonal with respect to the expectation $E_{\mathbf{U}}$. As a result,

$$\begin{aligned}
\mathrm{Var}[Q_n(\mathbf{X})|\mathbf{e},\mathbf{V}] &= E[A_n^2|\mathbf{e},\mathbf{V}] + E[B_n^2|\mathbf{e},\mathbf{V}] + \mathrm{Var}[C_n|\mathbf{e},\mathbf{V}] \\
&\geqslant E[A_n^2|\mathbf{e},\mathbf{V}] \\
&\geqslant \alpha^{2d}3^{-d}d! \sum_{i_1,\ldots,i_d=1}^{n} a_n(i_1,\ldots,i_d)^2 \varepsilon_{i_1}\ldots\varepsilon_{i_d}.
\end{aligned}$$

To go one step further, let us decompose ε_i into $(\varepsilon_i - p) + p$ and use (iv), so to obtain

$$\begin{aligned}
&\mathrm{Var}[Q_n(\mathbf{X})|\mathbf{e},\mathbf{V}] \\
\geq\ & d!\left(\frac{\alpha^2 p}{3}\right)^d + \alpha^{2d}3^{-d}d! \sum_{i_1,\cdots,i_d=1}^{n} a_n(i_1,\ldots,i_d)^2 \sum_{k=1}^{d} p^{d-k} \sum_{\{j_1,\cdots,j_k\}\subset\{1,\cdots,d\}} (\varepsilon_{i_{j_1}}-p)\cdots(\varepsilon_{i_{j_k}}-p) \\
=\ & d!\left(\frac{\alpha^2 p}{3}\right)^d + \alpha^{2d}3^{-d}d! \sum_{k=1}^{d} p^{d-k} \sum_{\{j_1,\cdots,j_k\}\subset\{1,\cdots,d\}} \sum_{i_1,\cdots,i_d=1}^{n} a_n(i_1,\ldots,i_d)^2 (\varepsilon_{i_{j_1}}-p)\cdots(\varepsilon_{i_{j_k}}-p) \\
=\ & d!\left(\frac{\alpha^2 p}{3}\right)^d + \alpha^{2d}3^{-d}d! \sum_{k=1}^{d} p^{d-k}\binom{d}{k} \sum_{i_1,\ldots,i_d=1}^{n} a_n(i_1,\ldots,i_d)^2 (\varepsilon_{i_1}-p)\ldots(\varepsilon_{i_k}-p).
\end{aligned}$$

where, in the last equality, we have used the symmetry (iii) of the coefficients a_n. Using (ii), (iv) and (v) we can write, for any fixed $k \in \{1, \ldots, d\}$,

$$
\begin{aligned}
& E\left[\left(\sum_{i_1,\ldots,i_d=1}^{n} a_n(i_1,\ldots,i_d)^2(\varepsilon_{i_1}-p)\ldots(\varepsilon_{i_k}-p)\right)^2\right] \\
&= p^k(1-p)^k \sum_{i_1,\ldots,i_k=1}^{n}\left(\sum_{i_{k+1},\ldots,i_d=1}^{n} a_n(i_1,\ldots,i_d)^2\right)^2 \\
&\leqslant \sum_{i_1=1}^{n}\left(\sum_{i_2,\ldots,i_d=1}^{n} a_n(i_1,\ldots,i_d)^2\right)^2 \\
&\leqslant \max_{1\leqslant j_1\leqslant n} \sum_{j_2,\ldots,j_d=1}^{n} a_n(j_1,\ldots,j_d)^2 \times \sum_{i_1,\ldots,i_d=1}^{n} a_n(i_1,\ldots,i_d)^2 \\
&= \frac{1}{d!} \max_{1\leqslant j_1\leqslant n} \sum_{j_2,\ldots,j_d=1}^{n} a_n(j_1,\ldots,j_d)^2 \ \to 0 \quad \text{as } n\to\infty.
\end{aligned}
$$

We deduce that, in probability,

$$
\liminf_{n\to\infty} \mathrm{Var}[Q_n(\mathbf{X})|\mathbf{e},\mathbf{V}] \geqslant d!\left(\frac{\alpha^2 p}{3}\right)^d. \tag{3.13}
$$

Convention. From now on, and since all the quantities we are dealing with are measurable with respect to $\mathbf{e}$, $\mathbf{U}$ and $\mathbf{V}$, we shall write $E_{\mathbf{U}}$ (resp. $E_{\mathbf{e},\mathbf{V}}$) to indicate the mathematical expectation with respect to $\mathbf{U}$ (resp. $\mathbf{e}$ and $\mathbf{V}$). Note that $E_{\mathbf{U}}$ coincides with the conditional expectation $E[\cdot|\mathbf{e},\mathbf{V}]$.

Step 2. Set $p_\eta(x) = \frac{1}{\eta\sqrt{2\pi}}e^{-\frac{x^2}{2\eta^2}}$, $x \in \mathbb{R}$, $0 < \eta \leqslant 1$, and let $\phi \in C_c^\infty$ be bounded by 1. It is immediately checked that

$$
\|\phi \star p_\eta\|_\infty \leqslant 1 \leqslant \frac{1}{\eta} \quad \text{and} \quad \|(\phi \star p_\eta)'\|_\infty \leqslant \frac{1}{\eta}. \tag{3.14}
$$

Let $d_{\mathrm{FM}}(X,Y) = \sup_{\phi:\,|\phi(x)-\phi(y)|\leqslant|x-y|} \big|E[\phi(X)] - E[\phi(Y)]\big|$ denote the Fortet-Mourier distance between (the distributions of) the two random variables X and Y; it is known that d_{FM} metrizes the convergence in law. We can write

$$
\begin{aligned}
& |E[\phi(Q_n(\mathbf{X})] - E[\phi(Q_n(\mathbf{Y})]| \\
\leqslant\ & |E[(\phi - \phi\star p_\eta)(Q_n(\mathbf{X}))]| + |E[(\phi - \phi\star p_\eta)(Q_n(\mathbf{Y}))]| + \frac{1}{\eta} d_{FM}(Q_n(\mathbf{X}), Q_n(\mathbf{Y})),
\end{aligned}
$$

Let us concentrate on the first two terms. We have, e.g., for the first term:

$$
\begin{aligned}
&|E[(\phi-\phi\star p_\eta)(Q_n(\mathbf{X}))]| \\
\leqslant\ & \left|E\left[(\phi-\phi\star p_\eta)(Q_n(\mathbf{X}))\mathbf{1}_{\{\mathrm{Var}[Q_n(\mathbf{X})|\mathbf{e},\mathbf{V}]<\frac{d!}{2}\left(\frac{\alpha^2 p}{3}\right)^d\}}\right]\right| \\
&+\left|E\left[(\phi-\phi\star p_\eta)(Q_n(\mathbf{X}))\mathbf{1}_{\{E[Q_n(\mathbf{X})^2|\mathbf{e},\mathbf{V}]>M\}}\right]\right| \\
&+\left|E\left[(\phi-\phi\star p_\eta)(Q_n(\mathbf{X}))\mathbf{1}_{\{\mathrm{Var}[Q_n(\mathbf{X})|\mathbf{e},\mathbf{V}]\geqslant\frac{d!}{2}\left(\frac{\alpha^2 p}{3}\right)^d,\ E[Q_n(\mathbf{X})^2|\mathbf{e},\mathbf{V}]\leqslant M\}}\right]\right| \\
\leqslant\ & 2P\left(\mathrm{Var}[Q_n(\mathbf{X})|\mathbf{e},\mathbf{V}]<\frac{d!}{2}\left(\frac{\alpha^2 p}{3}\right)^d\right)+2P\left(E[Q_n(\mathbf{X})^2|\mathbf{e},\mathbf{V}]>M\right) \\
&+\left|E\left[(\phi-\phi\star p_\eta)(Q_n(\mathbf{X}))\mathbf{1}_{\{\mathrm{Var}[Q_n(\mathbf{X})|\mathbf{e},\mathbf{V}]\geqslant\frac{d!}{2}\left(\frac{\alpha^2 p}{3}\right)^d,\ E[Q_n(\mathbf{X})^2|\mathbf{e},\mathbf{V}]\leqslant M\}}\right]\right|.
\end{aligned}
$$

We have, using the Markov inequality,

$$
P\left(E\left[Q_n(\mathbf{X})^2|\mathbf{e},\mathbf{V}\right]>M\right)\leqslant\frac{1}{M}E\left[E[Q_n(\mathbf{X})^2|\mathbf{e},\mathbf{V}]\right]=\frac{1}{M}.
$$

On the other hand,

$$
\begin{aligned}
&\left|E\left[(\phi-\phi\star p_\eta)(Q_n(\mathbf{X}))\mathbf{1}_{\{\mathrm{Var}[Q_n(\mathbf{X})|\mathbf{e},\mathbf{V}]\geqslant\frac{d!}{2}\left(\frac{\alpha^2 p}{3}\right)^d,\ E[Q_n(\mathbf{X})^2|\mathbf{e},\mathbf{V}]\leqslant M\}}\right]\right| \\
\leqslant\ & E_{\mathbf{e},\mathbf{V}}\left[\left|E_{\mathbf{U}}\left[(\phi-\phi\star p_\eta)\big(Q_n(\mathbf{X})\big)\right]\right|\mathbf{1}_{\{\mathrm{Var}[Q_n(\mathbf{X})|\mathbf{e},\mathbf{V}]\geqslant\frac{d!}{2}\left(\frac{\alpha^2 p}{3}\right)^d,\ E[Q_n(\mathbf{X})^2|\mathbf{e},\mathbf{V}]\leqslant M\}}\right].
\end{aligned}
$$

Therefore, in order to get our desired bound for $|E[\phi(Q_n(\mathbf{X})]-E[\phi(Q_n(\mathbf{Y})]|$, it only remains to analyze the term

$$
E_{\mathbf{e},\mathbf{V}}\left[\left|E_{\mathbf{U}}\left[(\phi-\phi\star p_\eta)\big(Q_n(\mathbf{X})\big)\right]\right|\mathbf{1}_{\{\mathrm{Var}[Q_n(\mathbf{X})|\mathbf{e},\mathbf{V}]\geqslant\frac{d!}{2}\left(\frac{\alpha^2 p}{3}\right)^d,\ E[Q_n(\mathbf{X})^2|\mathbf{e},\mathbf{V}]\leqslant M\}}\right],
$$

which is precisely the aim of the next steps.

Step 3. In this step, we shall introduce the framework we are going to use for the rest of the proof. We refer the reader to [1] for the details and missing proofs. Fix an integer m and let μ denote the distribution of the random vector $(X_1,\ldots,X_m)$, with $X_1,\ldots,X_m$ independent copies of $U\sim\mathcal{U}_{[-1,1]}$, There exists a reversible Markov process on $\mathbb{R}^m$, with semigroup P_t, equilibrium measure μ and generator $\mathcal{L}$ given by

$$
\mathcal{L}f(x)=\sum_{i=1}^{m}\Big((1-x_i^2)\partial_{ii}f-2x_i\,\partial_i f\Big),\quad x\in[-1,1]^m. \tag{3.15}
$$

The operator $\mathcal{L}$ is selfadjoint and negative semidefinite. We define the carré du champ operator Γ as

$$
\Gamma(f,g)(x)=\frac{1}{2}\big(\mathcal{L}(fg)(x)-f(x)\mathcal{L}g(x)-g(x)\mathcal{L}f(x)\big)=\sum_{i=1}^{m}(1-x_i^2)\partial_i f(x)\partial_i g(x). \tag{3.16}
$$

When $f = g$ we simply write $\Gamma(f)$ instead of $\Gamma(f,f)$. An important property satisfied by Γ is that it is diffusive in the following sense:

$$\Gamma(\phi(f), g) = \phi'(f)\Gamma(f,g). \tag{3.17}$$

Besides, see, e.g., [1, Sect. 2.7.4], the eigenvalues of $-\mathcal{L}$ are given by

$$\mathrm{Sp}(-\mathcal{L}) = \{i_1(i_1-1) + \cdots + i_m(i_m-1) \,|\, i_1, \ldots, i_m \in \mathbb{N}\}.$$

They may be ordered as a countable sequence like $0 = \lambda_0 < \lambda_1 < \lambda_2 < \cdots$, with a corresponding sequence of orthonormal eigenfunctions u_0, u_1, u_2, $\cdots$ where $u_0 = 1$; in addition, this sequence of eigenfunctions forms a complete orthogonal basis of $L^2(\mu)$. Also, note that the first nonzero element of $\mathrm{Sp}(-\mathcal{L})$ is $\lambda_1 = 1 > 0$. Also, one can compute that, when $\lambda \in \mathrm{Sp}(-\mathcal{L})$, then $\mathrm{Ker}(\mathcal{L}+\lambda I)$ is composed of those polynomial functions $R(x_1, \ldots, x_m)$ having the form

$$R(x_1, \ldots, x_m) = \sum_{i_1(i_1+1)+\cdots+i_m(i_m+1)=\lambda} \alpha(i_1, \cdots, i_{n_m}) J_{i_1}(x_1) \cdots J_{i_m}(x_m).$$

Here $J_i(X)$ is the ith Jacobi polynomial, defined as

$$J_i(x) = \frac{(-1)^i}{2^i i!} \frac{d^i}{dx^i} \left\{(1-x^2)^i\right\}, \quad x \in [-1,1].$$

To end up with this quick summary, we recal the following Poincaré inequality, that is immediate to prove by using the previous facts together with the decomposition $L^2(\mu) = \bigoplus_{\lambda \in \mathrm{Sp}(-\mathcal{L})} \mathrm{Ker}(\mathcal{L} + \lambda I)$:

$$\mathrm{Var}_\mu(f) \leqslant \int \Gamma(f) d\mu. \tag{3.18}$$

Step 4. We shall prove the existence of a constant $\kappa > 0$, depending on p, α and d but <u>not</u> on n, such that, for any $\delta > 0$,

$$\sup_{n \geqslant 1} \; E_{\mathbf{U}} \left[\frac{\delta}{\Gamma(Q_n)(\mathbf{X}) + \delta} \right] \mathbf{1}_{\{\mathrm{Var}[Q_n(\mathbf{X})|\mathbf{e},\mathbf{V}] \geqslant \frac{d!}{2} \left(\frac{\alpha^2 p}{3}\right)^d\}} \leqslant \kappa \, \delta^{\frac{1}{2d+1}}. \tag{3.19}$$

The proof of (3.19) will rely on the Poincaré inequality (3.18) which, here, takes the following form:

$$\mathrm{Var}[Q_n(\mathbf{X})|\mathbf{e}, \mathbf{V}] = \mathrm{Var}_{\mathbf{U}}[Q_n(\mathbf{X})] \leqslant E_{\mathbf{U}}[\Gamma(Q_n)(\mathbf{X})]. \tag{3.20}$$

Another ingredient is the Carbery-Wright inequality, that we recall for sake of completeness.

Theorem 3.1 *(see [9, Theorem 8]) There exists an absolute constant $c > 0$ such that, if $Q : \mathbb{R}^m \to \mathbb{R}$ is a polynomial of degree at most k and μ is a log-concave probability measure on $\mathbb{R}^m$, then, for all $\eta > 0$,*

$$\left(\int Q^2 d\mu \right)^{\frac{1}{2k}} \times \mu\{x \in \mathbb{R}^m : \, |Q(x)| \leqslant \eta\} \leqslant c\, k\, \eta^{\frac{1}{k}}. \tag{3.21}$$

Observe that the density of $\mathbf{U}$ is log-concave, as an indicator function of a convex set. Let us now proceed with the proof of (3.19). For any strictly positive u, and provided $\mathrm{Var}[Q_n(\mathbf{X})|\mathbf{e},\mathbf{V}] \geqslant \frac{d!}{2}\left(\frac{\alpha^2 p}{3}\right)^d$, one has

$$E\left[\frac{\delta}{\Gamma(Q_n)(\mathbf{X})+\delta}\right] \leqslant \frac{\delta}{u} + P(\Gamma(Q_n)(\mathbf{X}) \leqslant u) \leqslant \frac{\delta}{u} + c\, u^{\frac{1}{2d}}, \tag{3.22}$$

where $c > 0$ denotes a constant only depending on d, α and p and where the last inequality follows from the Carbery-Wright inequality (3.21), the inequality (3.20) and the fact that $\Gamma(Q_n)$ is a polynomial of order $2d$, see (3.16). Finally, choosing $u = \delta^{\frac{2d}{2d+1}}$ in (3.22) leads to the desired conclusion (3.19).

Step 5. We shall prove that

$$\sup_{n\geqslant 1}\left(E_{\mathbf{U}}[\Gamma(\Gamma(Q_n))(\mathbf{X})] + E_{\mathbf{U}}\left|(\mathcal{L}Q_n)(\mathbf{X})\right|\right)\mathbf{1}_{\{E[Q_n(\mathbf{X})^2|\mathbf{e},\mathbf{V}]\leqslant M\}} \leqslant c(M) \tag{3.23}$$

where $c(M)$ is a constant only depending on M (whose value may change from one line to another within this step). First, relying on the results of Step 3 we have that, for any n,

$$Q_n \in \bigoplus_{k\leqslant \lambda_{2d}} \mathrm{Ker}(\mathcal{L}+kI).$$

Since $\mathcal{L}$ is a bounded operator on the space $\bigoplus_{k\leqslant\lambda_{2d}}\mathrm{Ker}(\mathcal{L}+kI)$, we deduce immediately that $\sup_{n\geqslant 1} E_{\mathbf{U}}[(\mathcal{L}Q_n)(\mathbf{X})^2]\mathbf{1}_{\{E_{\mathbf{U}}[Q_n^2(\mathbf{X})]\leqslant M\}} \leqslant c(M)$. Besides, by the very definition (3.16) of Γ, one has $\Gamma(f,f) = \frac{1}{2}(\mathcal{L}+2\lambda I)(f^2)$ for $f \in \mathrm{Ker}(\mathcal{L}+\lambda I)$. Thus, one deduces for the same reason as above that

$$\sup_{n\geqslant 1} E_{\mathbf{U}}[\Gamma(\Gamma(Q_n))(\mathbf{X})]\mathbf{1}_{\{E_{\mathbf{U}}[Q_n^2(\mathbf{X})]\leqslant M\}} \leqslant c(M).$$

The proof of (3.23) is complete.

Step 6. We shall prove that, for any $n \geqslant 1$, any $0 < \eta \leqslant 1$, any $\delta > 0$ and any $M > 0$,

$$|E_{\mathbf{U}}[(\phi - \phi\star p_\eta)(Q_n(\mathbf{X}))]|\,\mathbf{1}_{\left\{\mathrm{Var}[Q_n(\mathbf{X})|\mathbf{e},\mathbf{V}]\geqslant\frac{d!}{2}\left(\frac{\alpha^2 p}{3}\right)^d, E[Q_n(\mathbf{X})^2|\mathbf{e},\mathbf{V}]\leqslant M\right\}} \leqslant 2\kappa\,\delta^{\frac{1}{2d+1}} + \sqrt{\frac{2}{\pi}}\,\frac{\eta}{\delta}\,c(M). \tag{3.24}$$

Using Step 4, one has

$$\begin{aligned}
&\left|E_{\mathbf{U}}\left[(\phi-\phi\star p_\eta)(Q_n(\mathbf{X}))\frac{\Gamma(Q_n)(\mathbf{X})+\delta}{\Gamma(Q_n)(\mathbf{X})+\delta}\right]\right|\mathbf{1}_{\left\{\mathrm{Var}[Q_n(\mathbf{X})|\mathbf{e},\mathbf{V}]\geqslant\frac{d!}{2}\left(\frac{\alpha^2 p}{3}\right)^d, E[Q_n(\mathbf{X})^2|\mathbf{e},\mathbf{V}]\leqslant M\right\}}\\
\leqslant\;& 2\,E_{\mathbf{U}}\left[\frac{\delta}{\Gamma(Q_n)(\mathbf{X})+\delta}\right]\mathbf{1}_{\left\{\mathrm{Var}[Q_n(\mathbf{X})|\mathbf{e},\mathbf{V}]\geqslant\frac{d!}{2}\left(\frac{\alpha^2 p}{3}\right)^d\right\}}\\
&+\left|E_{\mathbf{U}}\left[(\phi-\phi\star p_\eta)(Q_n(\mathbf{X}))\frac{\Gamma(Q_n)(\mathbf{X})}{\Gamma(Q_n)(\mathbf{X})+\delta}\right]\right|\mathbf{1}_{\{E[Q_n(\mathbf{X})^2|\mathbf{e},\mathbf{V}]\leqslant M\}}\\
\leqslant\;& 2\,\kappa\,\delta^{\frac{1}{2d+1}} + \left|E_{\mathbf{U}}\left[(\phi-\phi\star p_\eta)(Q_n(\mathbf{X}))\frac{\Gamma(Q_n)(\mathbf{X})}{\Gamma(Q_n)(\mathbf{X})+\delta}\right]\right|\mathbf{1}_{\{E[Q_n(\mathbf{X})^2|\mathbf{e},\mathbf{V}]\leqslant M\}}.
\end{aligned} \tag{3.25}$$

Now, set $\Psi(x)=\int_{-\infty}^{x}\phi(s)ds$ and let us apply (3.17). We obtain

$$\begin{aligned}
&\left|E_{\mathbf{U}}\left[(\phi-\phi\star p_\eta)(Q_n(\mathbf{X}))\frac{\Gamma(Q_n)(\mathbf{X})}{\Gamma(Q_n)(\mathbf{X})+\delta}\right]\right|\\
=&\ \left|E_{\mathbf{U}}\left[\frac{1}{\Gamma(Q_n)(\mathbf{X})+\delta}\,\Gamma\big((\Psi-\Psi\star p_\eta)\circ Q_n,Q_n\big)(\mathbf{X})\right]\right|\\
=&\ \left|E_{\mathbf{U}}\left[((\Psi-\Psi\star p_\eta)\circ Q_n)(\mathbf{X})\left(\Gamma\big(Q_n,\frac{1}{\Gamma(Q_n)+\delta}\big)(\mathbf{X})+\frac{(\mathcal{L}Q_n)(\mathbf{X})}{\Gamma(Q_n)(\mathbf{X})+\delta}\right)\right]\right|\\
=&\ \left|E_{\mathbf{U}}\left[((\Psi-\Psi\star p_\eta)\circ Q_n)(\mathbf{X})\left(-\frac{\Gamma(Q_n,\Gamma(Q_n))(\mathbf{X})}{(\Gamma(Q_n)(\mathbf{X})+\delta)^2}+\frac{(\mathcal{L}Q_n)(\mathbf{X})}{\Gamma(Q_n)(\mathbf{X})+\delta}\right)\right]\right|\\
\leqslant&\ \frac{1}{\delta}E_{\mathbf{U}}\left\{|((\Psi-\Psi\star p_\eta)\circ Q_n)(\mathbf{X})|\times\big[\Gamma(\Gamma(Q_n))(\mathbf{X})+\big|(\mathcal{L}Q_n)(\mathbf{X})\big|\big]\right\}. \qquad (3.26)
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
|\Psi(x)-\Psi\star p_\eta(x)| &= \left|\int_{\mathbb{R}}p_\eta(y)\left(\int_{-\infty}^{x}(\phi(u)-\phi(u-y))\,du\right)dy\right|\\
&\leqslant \int_{\mathbb{R}}p_\eta(y)\left|\int_{-\infty}^{x}\phi(u)du-\int_{-\infty}^{x}\phi(u-y)du\right|dy\\
&\leqslant \int_{\mathbb{R}}p_\eta(y)\left|\int_{x-y}^{x}\phi(u)du\right|dy\leqslant\int_{\mathbb{R}}p_\eta(y)\,|y|\,dy\leqslant\sqrt{\frac{2}{\pi}}\eta. \qquad (3.27)
\end{aligned}$$

The desired conclusion (3.24) now follows easily from (3.23), (3.25), (3.26) and (3.27).

Step 7: Concluding the proof. Combining the results of all the previous steps, we obtain, for any $n\geqslant 1$, any $0<\eta\leqslant 1$, any $\delta>0$ and any $M>0$,

$$\begin{aligned}
&\sup_{\phi\in C_c^\infty:\,\|\phi\|_\infty\leqslant 1}|E[\phi(Q_n(\mathbf{X})]-E[\phi(Q_n(\mathbf{Y})]|\\
\leqslant&\ \frac{1}{\eta}d_{FM}(Q_n(\mathbf{X}),Q_n(\mathbf{Y}))+\frac{4}{M}+2P\left(\mathrm{Var}[Q_n(\mathbf{X})|\mathbf{e},\mathbf{V}]<\frac{d!}{2}\left(\frac{\alpha^2p}{3}\right)^d\right) \qquad (3.28)\\
&+2P\left(\mathrm{Var}[Q_n(\mathbf{Y})|\mathbf{e},\mathbf{V}]<\frac{d!}{2}\left(\frac{\alpha^2p}{3}\right)^d\right)+4\kappa\,\delta^{\frac{1}{2d+1}}+2\sqrt{\frac{2}{\pi}}\,\frac{\eta}{\delta}\,c(M). \qquad (3.29)
\end{aligned}$$

In (3.28)-(3.29), take the limit $n\to\infty$. Due to (3.13) on one hand and Theorem 1.3 on the other hand (plus the fact that the Fortet-Mourier distance d_{FM} metrizes the convergence in distribution), one obtains

$$\limsup_{n\to\infty}\sup_{\phi\in C_c^\infty:\,\|\phi\|_\infty\leqslant 1}|E[\phi(Q_n(\mathbf{X})]-E[\phi(Q_n(\mathbf{Y})]|\leqslant\frac{4}{M}+4\kappa\,\delta^{\frac{1}{2d+1}}+2\sqrt{\frac{2}{\pi}}\,\frac{\eta}{\delta}\,c(M).$$

The desired conclusion (1.6) then follows by letting (in this order) $\eta\to 0$, $\delta\to 0$ and $M\to\infty$. ∎

Acknowledgments. We would like to express our gratitude to the referee for his/her careful reading of the manuscript and for his/her several suggestions.

References

[1] D. Bakry, I. Gentil and M. Ledoux (2014). *Analysis and Geometry of Markov Diffusion Semigroups.* Grundlehren der Mathematischen Wissenschaften **348**. Springer.

[2] A.R. Barron (1986). Entropy and the central limit theorem. *Ann. Probab.* **14**, no. 1, pp. 336-342.

[3] S. G. Bobkov, G. P. Chistyakov and F. Götze (2014): Berry-Esseen bounds in the entropic central limit theorem. *Probab. Theory Related Fields* **159**, no. 3-4, pp 435-478.

[4] S. G. Bobkov, G. P. Chistyakov and F. Götze (2014): Fisher information and the central limit theorem. *Probab. Theory Related Fields* **159**, no. 1-2, pp. 1-59

[5] S. G. Bobkov, G. P. Chistyakov and F. Götze (2013): Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem. *Ann. Probab.* **41**, no. 4, pp. 2479-2512.

[6] Borell, C. (1988). Real polynomial chaos and absolute continuity. *Probab. Theory Related Fields* **77**, no. 3, 397-400.

[7] T. Cacoullos, V. Papathanasiou, and S. A. Utev. (1994): Variational inequalities with examples and an application to the central limit theorem. *Ann. Probab.* **22**, no. 3, pp. 1607-1618.

[8] T. Cacoullos, V. Papathanasiou, and S. A. Utev. (1993): Another characterization of the normal law and a proof of the central limit theorem connected with it. *Theory of Probability and Its Applications* **37**, no. 4, pp. 581-588.

[9] A. Carbery and J. Wright (2001): Distributional and L^q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$. *Math. Research Lett.* **8**, pp. 233-248.

[10] B.V. Gnedenko and A.N. Kolmogorov (1968): *Limit distributions for sums of independent random variables.* Revised edition Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills., Ont. 1968 ix+293 pp.

[11] O.T. Johnson and A.R. Barron (2004): Fisher information inequalities and the central limit theorem. *Probab. Theory Related Fields* **129**, no.3, pp. 391-409.

[12] M. Mamatov and M.K. Halikov (1964): Global limit theorems for distribution functions in the higher-dimensional case. *Izv. Akad. Nauk UzSSR Ser. Fiz.-Mat. Nauk* no. 1, 13-21 (in Russian).

[13] I. Nourdin and G. Peccati (2012). *Normal Approximations with Malliavin Calculus. From Stein's Method to Universality.* Cambridge University Press.

[14] Nourdin, I., Peccati, G. and Swan, Y. (2013). Entropy and the fourth moment phenomenon. *Journal of Functional Analysis,* **266**, Issue 5, Pages 3170–3207.

[15] S.H. Sirazdinov and M. Mamatov (1962): On mean convergence for densities. *Teor. Verojatnost. i Primenen.* **7**, pp. 433-437 (in Russian).

[16] E. Mossel, R. O'Donnell and K. Oleszkiewicz (2010). Noise stability of functions with low influences: Variance and optimality. *Ann. Math.* **171**, pp. 295-341.

[17] I. Nourdin, G. Peccati, G. Poly and R. Simone (2014). An invariance principle under the total variation distance. *J. Theoret. Probab.*, to appear.

[18] I. Nourdin and G. Poly (2013). Convergence in total variation on Wiener chaos. *Stoch. Proc. Appl.* **123**, pp. 651-674.

[19] Y. Prohorov (1952): On a local limit theorem for densities. *Doklady Akad. Nauk SSSR (N.S.)* **83**, pp. 797-800. In Russian.

[20] V. I. Rotar (1979) Limit theorems for polynear forms. *J. Multiv. Anal.* **9, 511-530**

[21] R. Shimizu (1975): On Fisher's amount of information for location family. *In G.P.Patil et al (ed.), Statistical Distributions in Scientific Work* **3**, pp. 305-312. Reidel.